To cite this document:

Tuni, A., Rentizelas, A. and Duffy, A. (2018), "Environmental performance measurement for green supply chains: a systematic analysis and review of quantitative methods", *International Journal of Physical Distribution & Logistics Management*, doi: 10.1108/IJPDLM-02-2017-0062

Environmental performance measurement for green supply chains: a systematic analysis and review of quantitative methods

Abstract

Purpose – The majority of the environmental impacts in a typical supply chain can arise beyond the focal firm boundaries. However, no standardised method to quantify these impacts at the supply chain level currently exists. The aim of this work is to identify the quantitative methods developed to measure the environmental performance of supply chains and evaluate their key features.
Design/methodology/approach – A systematic literature review is conducted at the intersection of performance measurement and green supply chain management fields, covering 78 publications in peer-reviewed academic journals. The literature is reviewed according to several perspectives, including the environmental aspects considered, the main purpose of measurement, model types and the extent of supply chain covered by performance measurements.
Findings – Adopted environmental metrics show a low degree of standardisation and focus on natural resources, energy and emissions to air. The visibility and traceability of environmental aspects are still limited: the assessment of environmental impacts does not span in most cases beyond the direct business partners of the focal firms. A trade-off was observed between the range of environmental aspects and the extent of the supply chain considered with no method suitable for a holistic evaluation of the environmental supply chain performance identified. Three major streams of research developing in the field are identified, based on different scope.
Originality/value – This paper is the first attempt to examine in detail what tiers of the supply chain are actually involved in green performance assessment, ultimately contributing to clarify the scope of the supply chain dimension in green supply chain management performance measurement research. The work also recognises which methods are applicable to extended supply chains and explores how different methodologies perform in terms of supply chain extent covered.
Keywords Environmental metrics, Environmental performance measurement, Green supply chain management, Multi-tier supply chain, Literature review, Quantitative methods, Sustainability
Article classification Literature review

Introduction

Climate change and other environmental issues have received significant attention lately, capturing the interest of policy makers due to the potential impacts on both economies and people (Bloemhof et al., 2015). Since the Kyoto Protocol adoption in 1997, governments have been introducing stricter rules to control emissions in specific sectors and to limit the overall impact arising from industrial activities (Björklund et al., 2012). Organisations have lately been facing increased pressure from customers as well, demanding more sustainable products and services (Kovács, 2008). This pressure is enhanced by other stakeholders, such as non-governmental organisations (NGOs) and local communities, who are pressing for transparency and adequate reporting about the companies' activities and can damage their image (Meixell and Luoma, 2015).

These pressures initially targeted the focal firms, that are those organisations having a favourable balance of power in the supply chain, either by ruling the network or by providing the contact to the market (Seuring and Müller, 2008). The pressures later expanded to the supply network of focal firms due to a number of factors. Supply chains have lately become global and outsourcing processes to countries with low production cost, often coupled with looser environmental regulations and standards, has become a regular practice (Hutchins and Sutherland, 2008). Additionally, a significant ratio of the overall supply chain environmental impact arises in the extended supply chain of products, beyond the focal firm, including the usage phase and the end-of-life management (Veleva et al., 2003). The World Resource Institute (2009) concluded that companies beyond the focal firm are responsible for up to 80% of the overall supply chain emissions, with the extreme example of Marks & Spencer estimating its supply chain environmental impact to 90%, with only 10% attributed to the focal firm (Beavis, 2015). It is therefore evident that environmental performance cannot be adequately addressed at a single company level anymore; on the contrary, a holistic approach is needed, encompassing the whole supply chain (Fabbe-Costes et al., 2011; McIntyre et al., 1998).

Scholars have developed a variety of methods to measure the environmental performance at the supply chain level, addressing different environmental aspects and being applicable to various supply chain extents. However, a standardised method has not been established in the literature. The aim of this work is thus to identify quantitative methods developed to assess the environmental performance of supply chains, classify and evaluate their key features by systematically reviewing the literature at the intersection of the performance measurement and green supply chain management (GSCM) fields. This work is the first to identify which tiers of the supply chain are measured with respect to the environmental performance, in order to determine the suitability of different methods for the environmental performance measurement of extended supply chains. It also explores the features of the supply chains measured, such as the type of supply chain and whether a cradle-to-gate or a cradle-to-grave approach is adopted.

Background

Performance measurement in GSCM

The scope of this review lies at the intersection of two disciplines, performance measurement and GSCM. The former can be defined as the process of evaluating the efficiency and/or the effectiveness of an action (Neely et al., 1995). This is herein addressed specifically within the GSCM field, which can be defined as "integrating environmental thinking into supply chain management, including product design, material sourcing and selection, manufacturing processes, delivery of the final product to the consumers as well as end-of-life management of the product after its useful life" (Srivastava, 2007).

Traditionally, performance measurement targeted the economic dimension only, but increased competition on non-financial aspects forced companies to include other factors (Taticchi et al., 2013). Consequently, measuring environmental performance has become increasingly popular as its strategic role was recognised: green management was proven to enhance the long term competitiveness and economic performance (Rao and Holt, 2005). However, as competition shifted from a company-versus-company to a supply chain-versus-supply chain form, measuring performance, including environmental performance, cannot be executed at the single company level anymore, but requires a holistic approach encompassing the supply chain (Cabral et al., 2012). Extending the assessment to the supply chain is thus required to obtain a realistic evaluation of the environmental impacts of products.

The need to develop tools suitable for monitoring GSCM performance has also been evident in industry. Several organisations developed in-house methods to assess their direct suppliers' sustainability performance through scorecards (Renewable Choice Energy, 2012), while the SCOR framework, a framework for assessing supply chain processes performance, recently introduced a specific section on environmental aspects in version 11 (APICS, 2014).

Previous Relevant Literature Reviews

To provide the context to this work, previous relevant literature reviews were analysed[1]. Initially, literature reviews in the field of GSCM, such as Srivastava (2007) and Seuring and Müller (2008), addressed the problem context and set the theoretical grounds of the field.

The investigation of quantitative modelling and performance measurement for GSCM followed chronologically, with the exception of an early review of Hervani et al. (2005). A better definition of the purpose of the measurement (Björklund et al., 2012), a clearer definition of the meaning of sustainability performance and its related goals (Beske-Janssen et al., 2015), as well as a wider inclusion of stakeholders (Björklund et al., 2012) were identified as areas for improvement at the design stage of the performance measurement systems (PMS)

Considering the environmental impacts, reviews highlighted that, when multiple sustainability dimensions are addressed, methods tailored for supply chains very often only address a single environmental aspect (Beske-Janssen et al., 2015; Seuring, 2013; Taticchi et al., 2013, 2014). Limiting the performance metrics assessed when multiple organisations are involved may facilitate data collection and its use for decision support, whereas expanding the scope to multiple environmental impacts may challenge the methods' applicability (Liu et al., 2011).

Another hurdle to the diffusion of GSCM PMS was identified in the lack of standardisation in the field: despite SCOR gaining popularity in supply chain performance assessment, a universally accepted framework does not currently exist (Taticchi et al., 2013, 2014). Moreover, a significant number of metrics were developed in sustainable supply chain management (SSCM) with a low frequency repetition (Ahi and Searcy, 2015; Shaw et al., 2010).

Finally, despite the focus on quantitative environmental performance in supply chains, only three reviews investigated the features of supply chains involved. Beske-Janssen et al. (2015) pointed out

[1] Details about the main features of past reviews are available from the authors upon request

that the focal firm is usually the focus of the measurement despite the supply chain perspective being investigated, but also identified a surprising vagueness in the definition of the scope of the measurement in terms of supply chain tiers considered. Brandenburg et al. (2014) and Miemczyk et al. (2012) investigated the level of analysis of the proposed methods, proposing different classifications. However, despite the declared focus on GSCM, no review clearly investigated the supply chain extent covered by environmental performance measurement.

Thus, it can be concluded that existing reviews focused on the theoretical background of GSCM performance measurement, but lacked in providing an in-depth investigation of the supply chains the performance is measured across. Specifically, none of the previous reviews clearly investigates the supply chain extent covered by environmental performance measurement in conjunction with the type of supply chain, differentiating between forward, reverse and closed loop supply chains. This work addresses the knowledge gap identified above by focusing on the supply chain extent and other supply chain-related characteristics of methods developed for GSCM performance measurement. Moreover, it is the first work to evaluate the supply chain extent coverage of methods in relation to other key features of the methods, including environmental aspects considered, the purpose of measurement and model types, identifying relationships between them.

Methodology

A systematic process is adopted in this study, allowing a transparent and structured approach to investigate the body of knowledge in a specific field (Fink, 1998). Systematic literature reviews are widely accepted as a standardised approach to analyse published materials in the management field and are recognised for minimising bias in paper selection and offering the opportunity for research replication (Tranfield et al., 2003). The process followed to conduct the systematic literature review (Figure 1) is based on Jesson et al. (2011) but applies the inclusion and exclusion criteria stage twice instead of once.

Firstly, the need for the review was established through an iterative process of analysis of the literature in the field, focusing on previous related reviews. Secondly, the aim and scope were established, leading to three key research questions:

RQ1: What environmental performance metrics are adopted at the supply chain level?

RQ2: What extent of the supply chain, both upstream and downstream from the focal firm, are environmental performance measurement methods and related metrics addressing?

RQ3: What are the quantitative methods adopted to measure the environmental performance of supply chains? Is there a relationship between the type of method and the extent of supply chain covered or the scope of the work?

Two databases were selected for sourcing the articles. Scopus, as the largest peer-reviewed journal database in the management and engineering fields (Ahi and Searcy, 2013), and Web of Science, which particularly focuses on management (Taticchi et al., 2014). A structured combination of keywords was selected to conduct the database search (Figure 2). The four groups of terms encapsulate the review scope by highlighting the required supply chain context and focus on environmental dimensions (first two groups), whereas the last two groups target the environmental performance measurement. Published articles in English language in peer-reviewed journals up to 2015 were included.

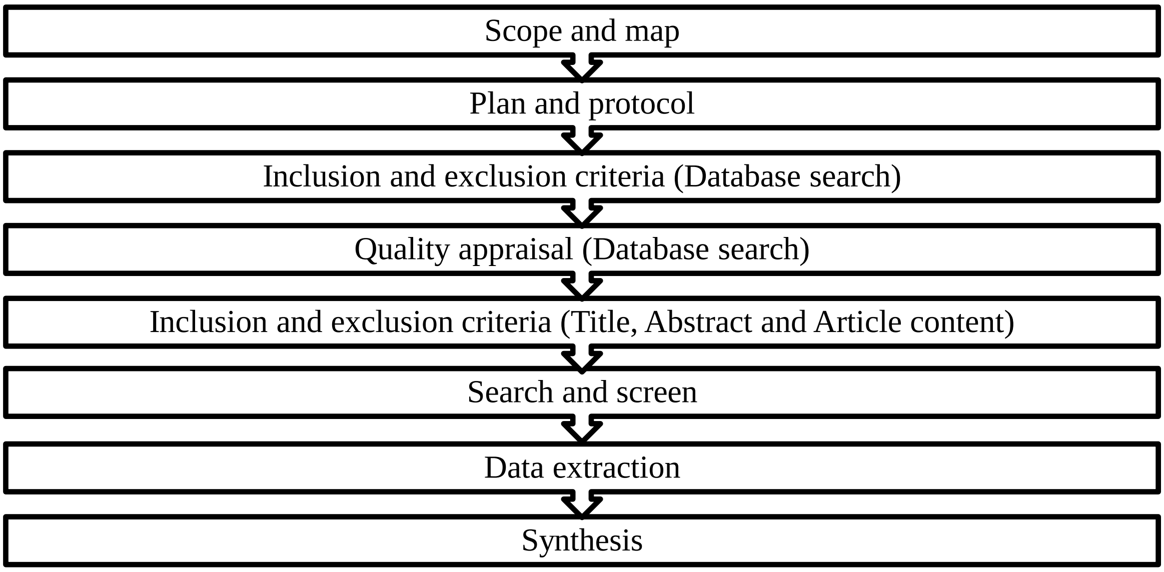

Fig.1: Systematic literature review process (based on Jesson et al., 2011)

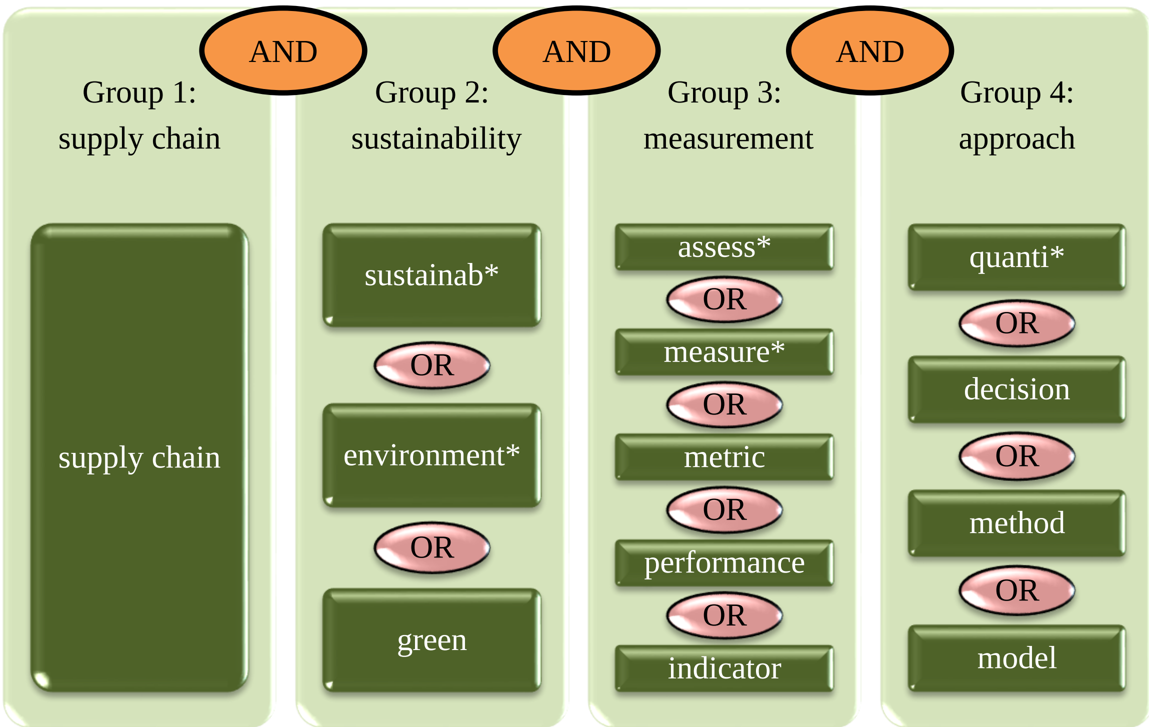

Fig. 2: Keywords used in the systematic review

The 4532 papers resulting from the keyword search went through a multiple stage-gate process. Duplicate papers were removed and article titles were screened for relevance, leaving 710 papers. Abstract screening followed, reducing the number to 185 papers. To increase research reliability, two reviewers performed these stages independently and compared the results. Any disagreement on paper inclusion was followed by discussion until final consensus was reached. Nine papers were not assessed due to full text unavailability. Finally, articles satisfying the inclusion criteria for both titles and abstract were read in full and went through the content analysis stage.

The criteria for inclusion at the content analysis stage were:

- Methodological dimension:

    - Explicit presentation of a method to assess environmental performance at the supply chain level. Applications or case studies only, without an explicit methodological contribution, were excluded;
    - Quantitative element in the methods should be explicit;
- Supply chain dimension:
    - Clear evidence of two or more tiers included in the environmental performance measurement;
    - Level of analysis limited to a single supply chain or single product. Papers with a wider level of analysis such as industrial network, industrial sectors and regional analysis were not considered;
- Environmental dimension: strong consideration of the environmental dimension of sustainability; the method should target the measurement of the environmental performance, rather than the enhancement and organisational efforts to achieve it.

After the full text screening, 78 papers were ultimately considered in the review. Each paper was analysed according to the following key aspects:

- Environmental performance: environmental inputs and outputs considered, distinct metrics adopted;
- Supply chain: number of tiers upstream and downstream of the focal firm involved in the environmental performance measurement; type of supply chain (forward, reverse, closed-loop); cradle-to-gate or cradle-to-grave approach;
- Methodology: model type, modelling technique and solution type;
- Scope of the work.

Bibliometric analysis

The temporal distribution of the 78 papers included in the analysis is depicted in Figure 3. The earliest publication is McIntyre et al. (1998), presenting the Environmental Performance Matrix to analyse the environmental performance of the Xerox supply chain. The chart indicates a steep increase in the published material starting from 2011 with the peak publications number reached in 2015 with 23 papers, indicating the novel and developing status of the research field.

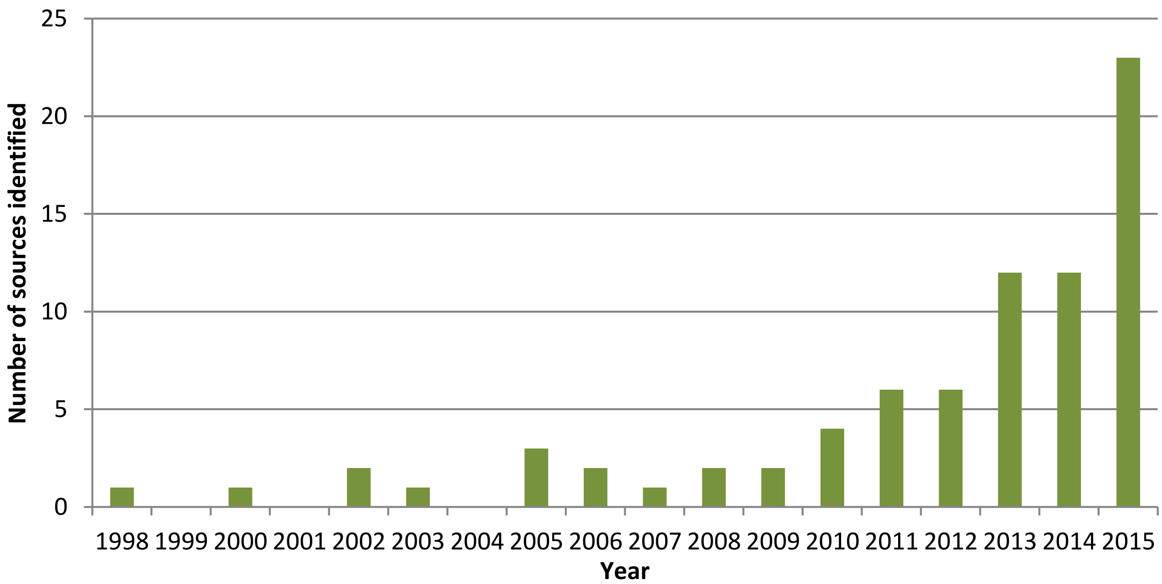

Fig. 3: Temporal distribution of papers

The 78 publications are spread over 42 journals, which can be clustered into two main groups: management-oriented journals and engineering journals with a more technical approach. Table I provides a summary of journals with multiple papers in the sample. Journal of Cleaner Production has the most publications (17 papers), accounting for 22% of the sample. The International Journal of Production Economics follows with 6 papers, whereas 5 publications appeared in International Journal of Production Research. Overall, almost 44% of publications appear in the top 5 journals in Table I. None of these journals has supply chain management as its core focus; it can thus be concluded that journals with an environmental or production focus are currently addressing the supply chain environmental performance topic. Finally, 30 journals appear only once in the sample, showing the multidisciplinary nature of the field: these journals cover various disciplines including mathematics, energy and computer science.

Table I
Distribution of papers by journal

Journal	Number of Articles	Authors[2]
Journal of Cleaner Production	17	(Baboulet and Lenzen, 2010; Brent and Visser, 2005; Govindan et al., 2013; Jakhar, 2015; Joa et al., 2014; Kannan et al., 2015, 2013; Kannegiesser and Günther, 2015; Lee, 2011; Manzardo et al., 2014; Mintcheva, 2005; Nagel, 2003; Nikolaou et al., 2013; Schmidt and Schwegler, 2008; Schmidt, 2015; Tajbakhsh and Hassini, 2015; Tsoulfas and Pappis, 2008)
International Journal of Production Economics	6	(Bai and Sarkis, 2010; Hashemi et al., 2015; Mahdiloo et al., 2015; Sarkis and Dhavale, 2015; Sundarakani et al., 2010; Zakeri et al., 2015)
International Journal of Production Research	5	(Azadnia et al., 2015; Brandenburg, 2015; Koh et al., 2012; Lu et al., 2007; Yakovleva et al., 2012)
Environmental Science and Technology	3	(Adhitya et al., 2011; Dewulf et al., 2005; De Soete, Debaveye, et al., 2014)
Resources, Conservation and Recycling	3	(Krikke, 2011; Shen et al., 2013; De Soete, Boone, et al., 2014)
ACS Sustainable Chemistry and Engineering	2	(Gao and You, 2015; Garcia and You, 2015)
Applied Energy	2	(Kravanja and Čuček, 2013; Rocco et al., 2014)
Ecological Indicators	2	(Alvarez and Rubio, 2015; Efroymson and Dale, 2015)
International Journal of Life Cycle Assessment	2	(Krikke, 2010; Röhrlich et al., 2000)
Production Planning & Control	2	(Dey and Cheffi, 2013; Tseng et al., 2013)
Supply Chain Management: An International Journal	2	(McIntyre et al., 1998; Varsei et al., 2014)

[2] The full reference list of reviewed papers is available from the authors upon request

Sustainability	2	(Salvado et al., 2015; Shokravi and Kurnia, 2014)
Other journals with a single paper	30	(Accorsi et al., 2015; Ahi and Searcy, 2014; Bernardi et al., 2012; Bojarski et al., 2009; Bouchery et al., 2012; Boukherroub et al., 2014; Caro et al., 2013; Charmondusit et al., 2014; De Soete et al., 2013; Dotoli et al., 2006; Du et al., 2015; Fahimnia et al., 2015; Gerbens-Leenes et al., 2003; Giarola et al., 2012; Jakhar, 2014; Jamshidi et al., 2012; Lee and Cheong, 2012; Mellor et al., 2002; Michelsen et al., 2006; Ortiz Gutiérrez et al., 2013; Pålsson et al., 2013; Ren et al., 2015, 2013; Shi et al., 2015; Trappey et al., 2012; Tuzkaya et al., 2009; Yazan et al., 2011; Yue et al., 2014; Zamboni et al., 2011; Zhang et al., 2014)

Results

Environmental aspects

The first investigated aspect is what type of environmental performance is measured, answering research question 1. The adopted classification has its methodological foundation in the transformation model by Slack et al. (2009), according to which each organisation in the supply network can be treated as a black box, taking into account only inputs and outputs. This approach is particularly suitable for the supply chain as it adopts a higher level of analysis, without investigating details within each organisation. The transformation model is adapted to analyse the environmental dimension of sustainability, following the well-established stream of research on the relationship between the economic and natural systems (UNEP, 2010). The classification of inputs and outputs categories follows Brent and Visser (2005), with two input and three output categories considered (Figure 4).

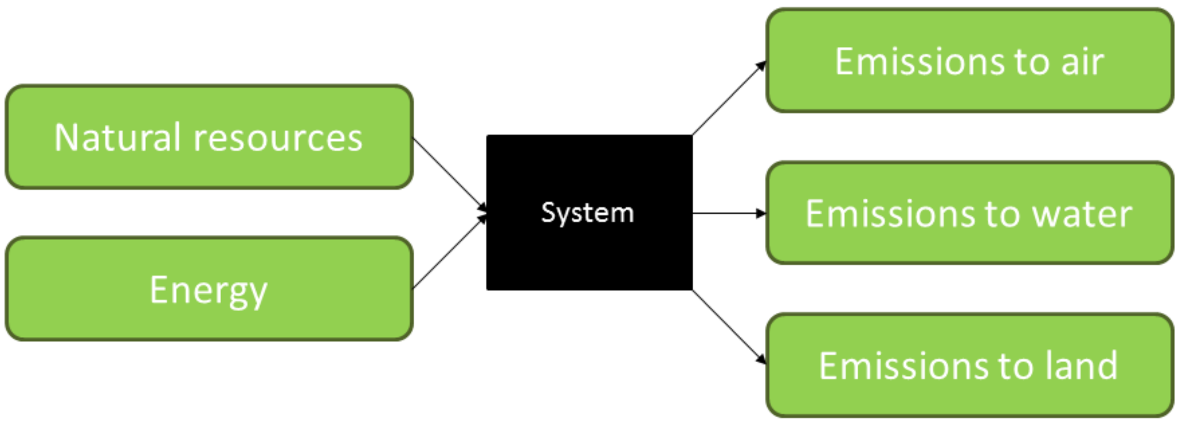

Fig. 4: Classification framework of environmental inputs and outputs

A. Frequency analysis of environmental measurements

The first step to understand what types of environmental performance are effectively considered by various methods is to identify the individual measurements adopted by each model and their positioning within the proposed classification framework. At the category level, “Natural Resources” was the most frequent one with 96 instances through 64 distinct measurements. Only 12% of papers included all inputs and outputs categories, therefore providing a complete coverage of the environmental dimension.

The specific metrics adopted were also recorded and grouped in thematic clusters to facilitate the analysis as shown in Table II, with each metric being assigned to a unique cluster. Metrics were assigned to clusters through keyword analysis, an established approach in the supply chain management field (Ahi and Searcy, 2013). Metrics that could not be assigned through this process were allocated to a cluster by similarity of scope. The authors performed the second step independently and reached consensus before the final allocation decision. The most frequent cluster is “Energy use”, with 37 instances through 13 distinct measurements. The clustering within each input and output category is discussed in detail in the following sub-section.

In the entire sample, 200 distinct measurements with 308 occurrences in total were identified, which equals exactly 4 environmental measurements considered on average in each paper. The ratio between the number of occurrences and the number of distinct metrics shows a very low repetition of metrics throughout the sample, in line with the observation of Ahi and Searcy (2015) in the wider context of SSCM. Metrics are often named differently despite conveying the same measurement (e.g. “Water consumption”, “Water usage”) or differ in being absolute (“Water use”) or relative (“Water use per unit of product”). Finally, some measurements are linked to targets (“Reduce the use of fresh water”).

Table II
Classification of environmental measurements

Environmental input or output	***Number of papers***	***Number of distinct/overall measurements***	***Measurement clusters***	***Number of distinct/overall measurements***	***Description of the cluster***
Environmental input					
Natural resources	*42*	*64/96*	*Water use*	*18/33*	*Use of water*
			Use of materials	*14/21*	*Use of raw or generic materials, without specific indication on their nature*
			Non-renewable resources consumption	*10/12*	*Use of materials and resources, including fossil resources, with a clear indication of their non-renewable nature*
			Use of recycled resources	*6/11*	*Use of resources originating from reverse supply chain activities.*
			Hazardous and Harmful materials use	*7/7*	*Use of dangerous materials classified as hazardous, toxic or harmful to humans*
			Land use	*4/7*	*Use of land*

			Use of packaging	*5/5*	*Use of packaging*
Energy	*41*	*23/47*	*Energy use*	*13/37*	*Use of energy from undefined sources*
			Renewable energy	*5/5*	*Explicit use of renewable sources of energy*
			Energy efficiency	*2/2*	*Efficiency in the use of energy*
			Other	*3/3*	*Energy metrics not falling under any of the above mentioned clusters*
Environmental output					
Emissions to air	*58*	*48/90*	*Carbon emissions*	*11/29*	*Emissions to air of polluting agents containing carbon, including CO, CO_2 and CH_4 emissions*
			GHG emissions	*5/19*	*Aggregate consideration of emissions from all greenhouse gases*
			Generic air emissions	*9/14*	*Undefined and generic emissions of polluting agents to air*
			Other specific air emissions	*10/13*	*Emissions to air of specified polluting agents, other than carbon emissions*
			Environmental impact related measurement	*11/13*	*Emissions classified under their ultimate environmental impact rather than on the basis of the emitted substances*
			Other	*2/2*	*Emissions to air metrics not falling under any of the above mentioned clusters*
Emissions to water	*12*	*15/16*	*Liquid waste*	*6/6*	*Undefined and generic liquid waste or spillage as well as effluents of specific liquid substances other than waste water*
			Waste water	*5/6*	*Waste water effluents*
			Environmental impact related measurement	*3/3*	*Emissions classified under their ultimate environmental impact rather than on the basis of the emitted substances*
			Other	*1/1*	*Emissions to water metrics not falling under any of the above mentioned clusters*
Emissions to land	*24*	*50/59*	*Solid waste produced*	*16/22*	*Undefined and generic solid waste as*

					well as emissions of specific solid substances to land
			Suitability for reverse chain (3Rs)	*15/17*	*Potential and/or effective use for recycling, reusing or remanufacturing activities of waste as well as any solid waste diverted from landfill*
			Hazardous and Harmful waste	*13/14*	*Solid waste, including toxic waste, requiring particular treatment due to the potential harm to humans*
			Other	*6/6*	*Emissions to land metrics not falling under any of the above mentioned clusters*
TOTAL		*200/308*			

B. Environmental inputs and outputs

Overall, 65% of the papers consider the environmental inputs: no significant preference was identified between the two inputs categories, as 42 papers consider resource consumption, whereas 41 incorporate energy use or consumption. Most addressed clusters include “Water use” for the natural resources category, whereas “Energy use” dominates the energy category. The majority of measurements adopted imply a negative correlation with the environmental impacts: an increase in input consumption leads to a worse environmental performance. The only exception is represented by renewable inputs, such as “Use of recycled resources” and “Renewable energy”.

83% of the papers consider environmental outputs. Unlike the environmental inputs case, scholars are mostly interested in one specific category, namely emissions to air, considered in 74% of the articles. On the other hand, emissions to land and water received less attention with 31% and 15% respectively. Most observed clusters include “Carbon emissions” in the emissions to air category, “Liquid waste” in the emissions to water category and “Solid waste produced” in the emissions to land category.

A number of reasons justify the identified extensive consideration of environmental inputs within the supply chain. Firstly, there is a need to consider resource consumption at a macro level, as “current levels of global production and consumption are using 50% more natural resources and services than ecosystems regenerate” (O’Rourke, 2014) and natural resource scarcity at the global level may even threaten the existence of certain supply chains (Bell et al., 2013). Secondly, it is impossible to reduce environmental outputs just by providing “end-of-pipe” solutions, but there is a need to reduce inputs proactively (McIntyre et al., 1998; De Soete et al., 2013). Although limiting the problem to an overall quantitative analysis without considering the mix and characteristics of inputs and outputs, Ritthof et al. (2002) reinforce this argument by stating that the pressure on the environment is automatically decreased if inputs are reduced, as they will inevitably become an output of the system at a certain point. Finally, reducing inputs is particularly attractive for organisations for economic reasons too, as they represent a cost. Therefore, such a reduction provides win-win opportunities involving both economic and environmental dimensions.

On the other hand, it is more common to find trade-off rather than win-win situations with the economic dimension in the case of environmental outputs: examples include Zhang et al. (2014), Boukherroub et al. (2014) and Mellor et al. (2002). Therefore, companies are less interested to evaluate their performance in terms of environmental outputs, when the monetary outcome is less tangible. Benefits arise in the longer term thanks to environmentally driven innovation and improved brand and image value, but are rarely visible in the short term (APICS, 2012). Air emissions are an exception to the output category as they are the single most addressed category in the sample considering both inputs and outputs. This interest could be attributed to regulatory schemes aiming to control carbon emissions introduced for different sectors in various geographical areas (Bouchery et al., 2012; Zakeri et al., 2015).

C. Contingency analysis of environmental categories

A contingency analysis of environmental categories was performed to identify association patterns between categories and pairs of categories whose combined observed frequency is higher or lower than the product of their single probabilities would suggest (Gold et al., 2010). The contingency analysis is performed through a chi-square test and calculated by the Phi-coefficient (ϕ), which identifies the patterns' strengths. While these patterns do not reveal causality and necessarily provide semantic argumentation, they provide statistical evidence that has to be justified (Gold et al., 2010). The contingency analysis was applied at the level of environmental categories, as the expected frequency of each pair needs to be bigger than five, a condition not achievable with a more detailed level of granularity (Fleiss, 1981).

Table III
Contingency results of environmental categories

Environmental categories pair		Expected frequency	Observed frequency	Chi-square significance	Phi coefficient
Energy	Emissions to land	12.8	22	0.000	0.518
Natural resources	Energy	22.4	32	0.000	0.504
Natural resources	Emissions to land	13.1	21	0.000	0.445
Natural resources	Emissions to water	6.5	12	0.001	0.392
Natural resources	Emissions to air	31.6	26	0.003	0.341
Energy	Emissions to air	30.9	26	0.010	0.295

As shown in Table III, three pairs show a Phi-coefficient above 0.4, which is considered the threshold of a strong association between the two categories, whereas three additional pairs fall in the range 0.2-0.4, which indicates moderate association (Cohen, 1969). Four pairs show a higher observed frequency than expected, showing a reinforcing association whereas the "natural resources – emissions to air" and "energy-emissions to air" pairs show a lower observed frequency than expected. While a justification for these pairs is found in the willingness of some authors to avoid double counting (Bojarski et al., 2009; Michelsen et al., 2006), this result stresses a less frequent application of "Emissions to air" in combination with other environmental categories. Indeed, "Emissions to air" are applied in isolation in 24 papers accounting for 31% of the sample.

Therefore, it can be concluded that emissions to air are often treated as a proxy of the overall environmental impact. Regulatory schemes played a significant role in this pattern. Focus on emissions by policy makers was prominent compared to other environmental impacts due to their direct effect on global warming (Pattara et al., 2012). This triggered the interest of academics to address managerial choices affecting environmental sustainability under different regulatory schemes, as in Bouchery et al. (2012), Caro et al. (2013), Fahimnia et al. (2015), and Zakeri et al. (2015).

On the other hand, “energy - emissions to land”, “natural resources - energy” and “natural resources - emissions to land” pairs have strong associations. Since waste-related clusters are dominant within the “emissions to land” category, it can be inferred that these associations identify strong relationship between those environmental categories that cause economic expenditure across the supply chain. As these categories are typically addressed simultaneously, they can be labelled as efficiency oriented, since the environmental performance improvement benefits the economic performance as well.

Supply Chain aspects

A. Supply Chain extent

The second aspect evaluated in this review is the extent of the supply chain effectively measured with respect to the environmental performance, answering research question 2. Previous reviews recognised that most environmental performance measurements for supply chains targeted a single organisation and its supply chain policies, rather than the supply chain (Brandenburg et al., 2014). However, a detailed mapping of which extent of the supply chain is covered by the current environmental measurement methods is still lacking in the literature.

In this work, a tier of the supply chain is defined as every individual organisation whose core activity is different from transportation activities only. In the case that transportation is integrated to other distribution services, such as warehousing, then this is considered as a separate tier of the supply chain for the purpose of this work. Vertically integrated supply chains with a number of activities taking place within the boundaries of a single firm are considered in this analysis as a single tier, even if activities occur in different geographical areas. The rationale behind this approach is that the decisions remain within the single organisation, eliminating challenges and barriers arising when multiple organisations are involved.

The extent of chains covered by environmental measurement is presented in Figure 5. The bars length represents the extent of the supply chain assessed, with the green part on the left representing the upstream network and the yellow part on the right representing the downstream network in respect to the focal firm. If more than three tiers either upstream or downstream are assessed, the method is considered suitable to evaluate the entire upstream/downstream network respectively. The bar width signifies the number of papers covering that specific case and corresponds to the respective circled number.

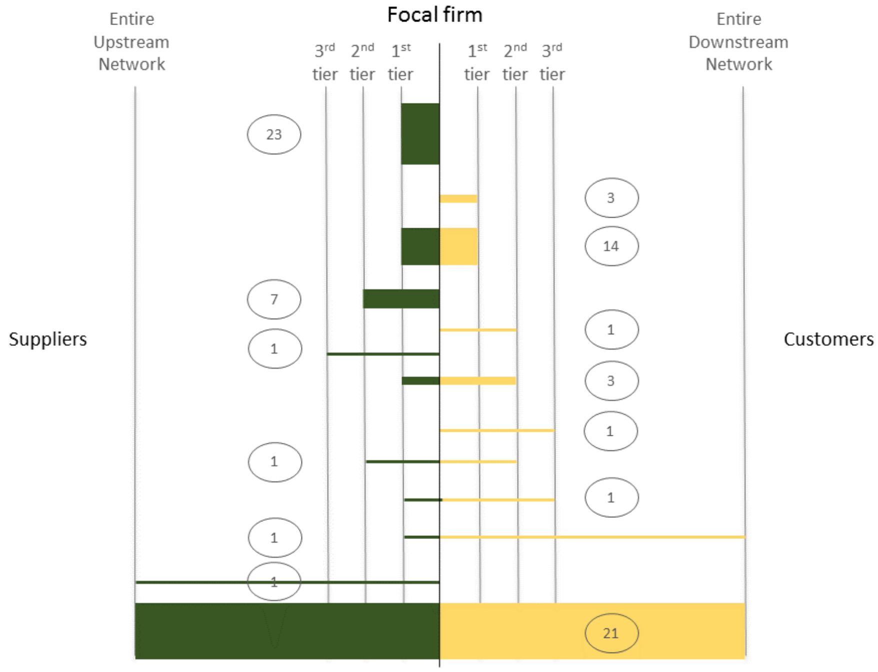

Fig. 5: Classification of methods based on the supply chain extent

Overall, 51% of the papers (40 in absolute figures) do not go beyond the first supply chain tier, upstream or downstream, limiting the supply chain extent covered to direct suppliers or customers only. This reflects a broader weakness of most companies in effectively mapping their supply chains, as "49 per cent of global manufacturing executives admit that their companies do not have supply chain visibility beyond their Tier 1 suppliers" (O'Rourke, 2014). This narrow approach is unable to evaluate accurately not only traditional economic aspects but sustainability as well: the extended supply chain needs to be fully assessed to obtain a complete sustainability profile (Miemczyk et al., 2012) as the biggest portion of environmental impacts arise in the extended supply chain (Veleva et al., 2003). The review shows that only 27% of papers (21) consider the entire upstream and downstream network, providing a full coverage of the supply chain.

An effective understanding of the entire supply chain is a prerequisite for measuring performance across multiple organisations. An accurate supply chain mapping can provide a solution if a centralised approach is adopted with the focal firm leading the measurement process thanks to its superior contractual power (Erol et al., 2011). An alternative decentralised approach is proposed by other authors, such as Schmidt and Schwegler (2008) or Nagel (2003) to overcome the rigidity of a centralised approach, especially in long and complex supply chains. However, a high level of standardisation of data is required to adopt a decentralised approach, which is yet to be achieved, as concluded in the "Environmental aspects" sub-section.

Despite over half of the articles (53%) targeting both upstream and downstream chains to obtain a comprehensive evaluation (Brent and Visser, 2005), the upstream network is more frequently addressed: 94% of papers (73) include at least one upstream chain tier, compared to 59% for downstream chain (46). This can originate from the fact that customers' reputation is influenced by the environmental reputation of their suppliers, whereas suppliers are not affected by the reputation of their customers (Kovács, 2008). Major scandals originating from inappropriate code of conduct of suppliers impacted several organisations, pushing them to assess the sustainability performance of their suppliers including both environmental and social aspects (Gimenez and Tachizawa, 2012).

However, focal firms are able to influence their suppliers' behaviour only when they have a reasonably favourable power balance along the chain, based on "purchasing power, duration of relation, personal relations and knowledge, reputation in the market" (Michelsen and Fet, 2010).

On the other hand, organisations have limited influence on the behaviour of the downstream part of the chain (Mentzer et al., 2001). Moreover, the position of the focal firm along the chain, when specified by authors, is typically standing midstream or downstream within the network, usually closer to the final customer. This can justify the lower number of papers addressing the downstream network, as it naturally limits the available number of tiers downstream compared to the number of tiers upstream.

B. Cradle-to-gate and cradle-to-grave approaches

A further analysis considered the type of supply chain addressed by methods included in the review combined with the cradle-to-gate or cradle-to-grave approach adopted (Table IV).

Three supply chains types are considered: the traditional forward supply chain, considering the material and information flow downstream from raw materials to the end customer (Stevens, 1989); the reverse chain, originating from the customers in the upstream direction (Nikolaou et al., 2013); the closed-loop supply chain, which is the combination of forward and reverse chains (Liu et al., 2011). Methods were found to target forward supply chains in 81% of the cases. Remaining papers address closed-loop supply chains, with the exceptions of Nikolaou et al. (2013) and Krikke (2011), who consider a reverse chain. The limited consideration of reverse chains indicates limited interest for them considered in isolation, whereas their inclusion in a closed-loop perspective together with the related forward chain looks more appealing to assess the overall benefit to the environment. Finally, Dotoli et al. (2006), Pålsson et al. (2013) and Trappey et al. (2012) consider both forward and closed-loop supply chains in their work, thus values in Table IV exceed the number of papers reviewed.

Table IV
Type of supply chain

	Cradle-to-Gate	**Cradle-to-Grave**	**Total**
Forward	52	14	66
Closed-loop	2	11	13
Reverse	2	0	2
Total	56	25	81

A cradle-to-gate approach considers all supply chain stages from raw material extraction up to the finished product (Ritthof et al., 2002). A cradle-to-grave scenario extends this view by adopting a lifecycle perspective, considering also the product usage phase and end-of-life management. When the product undergoes recycling, this approach is referred by some authors as cradle-to-cradle, as original materials re-enter a forward supply chain (Bloemhof et al., 2015). Cradle-to-gate scenarios naturally neglect part of the environmental impacts underestimating the overall environmental impact caused by products, especially in sectors where the direct impacts (Chatzinikolaou and Ventikos, 2015) and indirect impacts (Cichorowski et al., 2015) during the usage phase can have the most significant contribution.

Despite Elkington (2004) identifying over a decade ago a progressive change in the behaviour of companies towards an inclusive consideration of lifecycle stages following the point of sale, the identified methods rarely consider the product usage phase and end-of-life management in the performance measurement. This is particularly evident when forward supply chains are considered, where only 21% of the methods consider a cradle-to-grave scenario. The limited control of companies on the usage and end-of-life management stages as well as the difficulty in effectively measuring environmental performance during those stages can be considered among the main reasons limiting the adoption of cradle-to-grave approaches (Michelsen et al., 2006).

Data from Table IV show a strong association between forward supply chain and cradle-to-gate approach as well as between closed-loop supply chain and cradle-to-grave approach. This indicates that the supply chain evaluation is mostly focused on the pre-usage stages unless a lifecycle perspective is adopted. Issues about product responsibility in the usage phase are often neglected from the analysis of forward supply chains as well as the end-of-life treatments evaluation due to uncertainties about different end-of-life options (Michelsen et al., 2006). On the other hand, the lifecycle perspective is a common feature of closed-loop supply chains and cradle-to-grave approach. Recent regulations, such as the EU Waste Electrical and Electronic Equipment (WEEE) directive are trying to incorporate this extended perspective into regulatory schemes. A challenge still stands though for researchers to further incorporate the lifecycle perspective within effective supply chain environmental performance measurement tools.

Methodological approaches

In this section, papers are analysed based on the methodology they adopt to assess the environmental performance of supply chains to understand what are the leading approaches in GSCM performance measurement. The authors adapted the classification by Brandenburg et al. (2014), who evaluated quantitative models for supply chains under various perspectives. A number of additions to the classification scheme were required as some papers could not be accurately allocated to an existing category. The adopted classification scheme is presented in Figure 6, whereas Table V shows the model types and modelling techniques analysis.

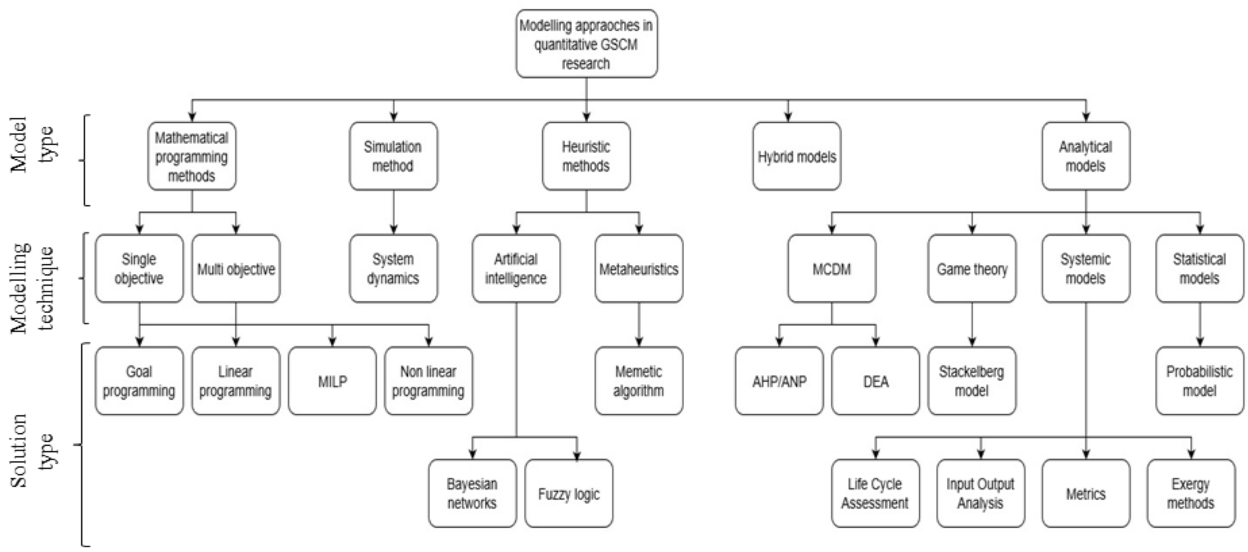

Fig. 6: Categories to evaluate quantitative methodological approaches for GSCM performance measurement (adapted from Brandenburg et al., 2014)

Table V
Classification of quantitative methodological approaches for GSCM performance measurement

Model type		Modelling technique		Solution approach	
Mathematical programming	21	Single objective	2	Goal programming	1
		Multi objective	20	Linear programming	2
				MILP	11
				Non-linear programming	7
Simulation	1	System dynamics	1		
Heuristic	8	Artificial intelligence	7	Bayesian networks	1
				Fuzzy logic	6
		Meta-heuristic	1	Memetic algorithm	1
Analytical	35	Game theory	2	Stackelberg model	1
				Unspecified	1
		MCDM	4	AHP/ANP	2
				DEA	2
		Statistical model	1	Probabilistic model	1
		Systemic model	25	Life Cycle Analysis	4
				Input / Output Analysis	3
				Metrics	13
				Exergy methods	5
		Multiple	3	AHP and Metrics	3
Hybrid	13	Other	13	Other	13

Analytical models are the dominant model type with 35 occurrences. Within this, systemic models are the most adopted modelling technique, followed by multi criteria decision-making (MCDM). The combination of both modelling techniques is common, with MCDM used to weight criteria based on opinion of stakeholders and decision makers in order to link the PMS to the supply chain strategy, while metrics are used to evaluate the environmental performance. Mathematical programming methods follow with 21 occurrences. The adopted modelling technique is always multi-objective in this case, with the single exception of Ren et al. (2015). Additionally, Dotoli et al. (2006) adopt both single and multi-objective modelling techniques. Heuristic methods are represented in 8 papers, whereas Adhitya et al. (2011) are the only authors adopting a simulation method, using system dynamics to evaluate the environmental performance of a diaper's supply chain. Finally, a common approach is using hybrid or multiple models within the same paper: this has been recognised as a way to overcome limitations of single methods (Saunders et al., 1997). Various combinations are frequently identified in the sample: the use of heuristic methods, especially fuzzy logic, is often combined with analytical models or mathematical programming methods to include uncertainty in the model, replicating more accurately conditions faced by organisations in their operations.

Relationship between supply chain extent and methodology

The relationship between the supply chain extent covered and the methodology adopted is explored, to analyse whether specific methodologies are more suitable to evaluate the environmental performance of particular supply chain configurations.

The supply chain extent configurations analysed earlier are clustered in four groups:

1. Dyad: either supplier-focal firm or focal firm-customer configuration

2. Triad: supplier-focal firm-customer
3. Multi-tier: configurations involving suppliers or customers beyond the 1st tier from the focal firm, but not including the entire network
4. Extended supply chain: entire upstream and downstream network;

Identifying relationships was not meaningful for simulation methods as only Adhitya et al. (2011) adopt such an approach. Therefore, the analysis considered only the four remaining model types. Table VI shows the occurrences of each model type against the supply chain extent configurations.

Table VI
Relationship between model type and supply chain extent

	Dyad	Triad	Other	Extended	Total
Mathematical programming	7	3	6	5	21
Heuristic methods	7	0	1	0	8
Hybrid methods	5	4	2	2	13
Analytical models	7	7	7	14	35
Total	26	14	16	21	77

Mathematical programming methods prove to be similarly adaptable to different supply chain configurations, with a peak for short dyadic supply chains. Hybrid methods are also applied for different supply chain configurations, with occurrences dropping when the extent of supply chain expands. Heuristic methods are used almost exclusively to address dyads: Jamshidi et al. (2012) are the only exception, trying to extend the evaluation of the supply chain beyond the direct suppliers. It can be thus concluded that the above model types are predominantly applied to dyads, limiting significantly the extent of supply chain effectively measured with respect to environmental performance.

On the other hand, analytical models are used in every supply chain configuration but show higher occurrences as the extent of the supply chain increases. Only 20% of analytical models target dyads, below the average of other model types, whereas 40% tackle extended supply chains, a significantly higher occurrence compared to other model types (Figure 7).

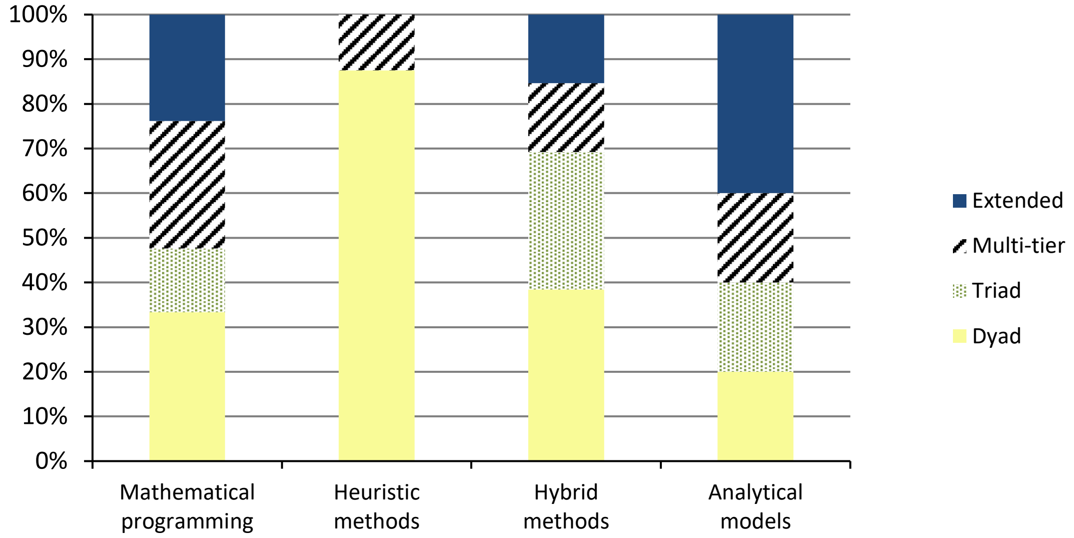

Fig. 7: Supply chain extent covered by model type

Analysing using the supply chain extent as a focal point (Figure 8), only the extended supply chain configuration has a clear direction in terms of model type use: 67% of papers with this configuration adopt analytical models. This result may support future researchers wanting to assess the environmental performance of extended supply chains, indicating systemic models or MCDM as the most frequently used modelling techniques.

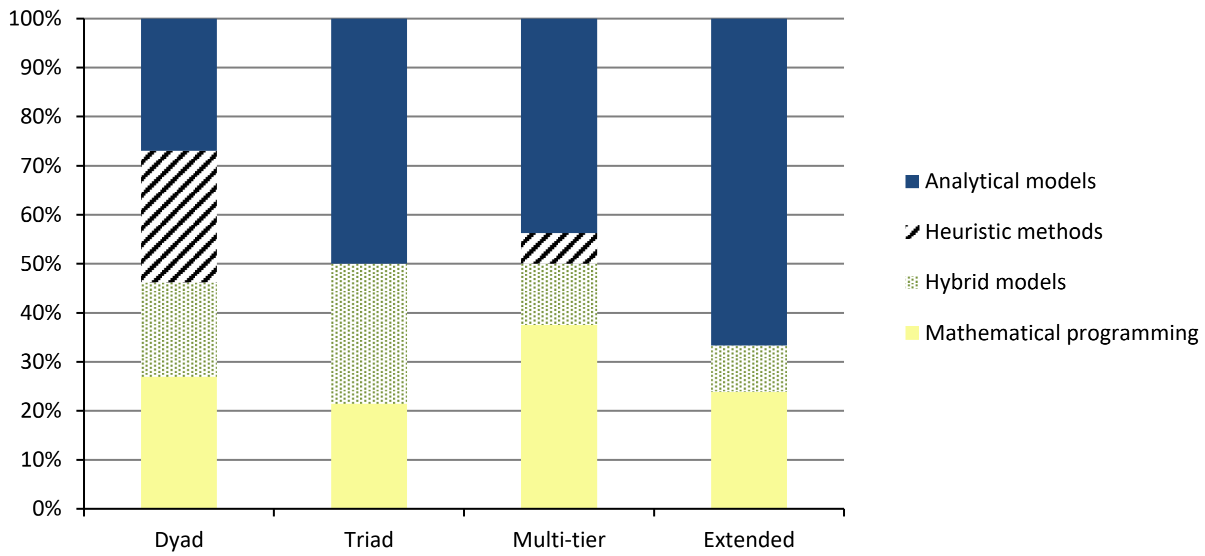

Fig. 8: Model types by supply chain extent

In order to examine the statistical significance of the above results, Cramer's V measure of association between model types and supply chain extent was utilised (Kateri, 2014), which is a Chi-square-based test (χ^2), specifically tailored for tables with dimensions higher than 2x2 and is calculated as (Liebtrau, 1983):

$$V = \sqrt{\frac{\varphi^2}{t}}$$

Where φ is the square root of χ^2divided by the number of total occurrences, and t is the minimum between the number of rows minus one and the number of columns minus one. Since a 4x4 square table is considered here, *t* equals 3. Based on Cohen's (1988) guidelines to interpret the Cramer's V results, for *t*=3, a small effect is associated to the value of 0.06, medium effect to 0.17 and large effect to 0.29. In the table under investigation Cramer's V is equal to 0.278. The test shows an approximate significance of 0.037, the results thus being statistically significant at the 5% level. Therefore, the contingency analysis indicates a significant effect relationship between the variables examined verifying a strong association between model types and supply chain extent.

Relationship between methodology and scope of the methods

This section introduces the final perspective of analysis, which is the primary scope of the papers. Three categories of scope were identified:

- Supply chain assessment (40 papers): the aim is to evaluate the supply chain performance from an environmental dimension only or along with the economic and/or social sustainability dimension.
- Supplier selection and evaluation (14 papers): the focus is on the process of evaluating and selecting suppliers, considering environmental criteria along with traditional criteria such as cost, quality and service level.
- Supply chain performance optimisation and supply chain design or re-design (24 papers): the purpose is to optimise the supply chain performance by considering multiple objectives, including the environmental impact. This involves decisions such as capacity assignments, flow allocation and mode of transportations in either greenfield or existing supply chains.

Figure 9 shows the relationship between the model type and the scope. Three strong associations are identified, showing a consensus among scholars in model types used to fit each scope. Mathematical programming is mostly used to optimise the performance or to design and plan the supply chain (in 86% of the cases), whereas 75% of papers with this scope adopt this method. Heuristic methods are primarily used to select and evaluate suppliers, with the only exception of Jamshidi et al. (2012). Despite representing just one tenth of the entire sample, heuristic methods constitute 50% of papers aiming to select and evaluate suppliers. Analytical models are mainly adopted for the assessment of the supply chain performance (in 89% of instances). Finally, hybrid models are applied with different scopes, reflecting the variety of methods adopted in this category.

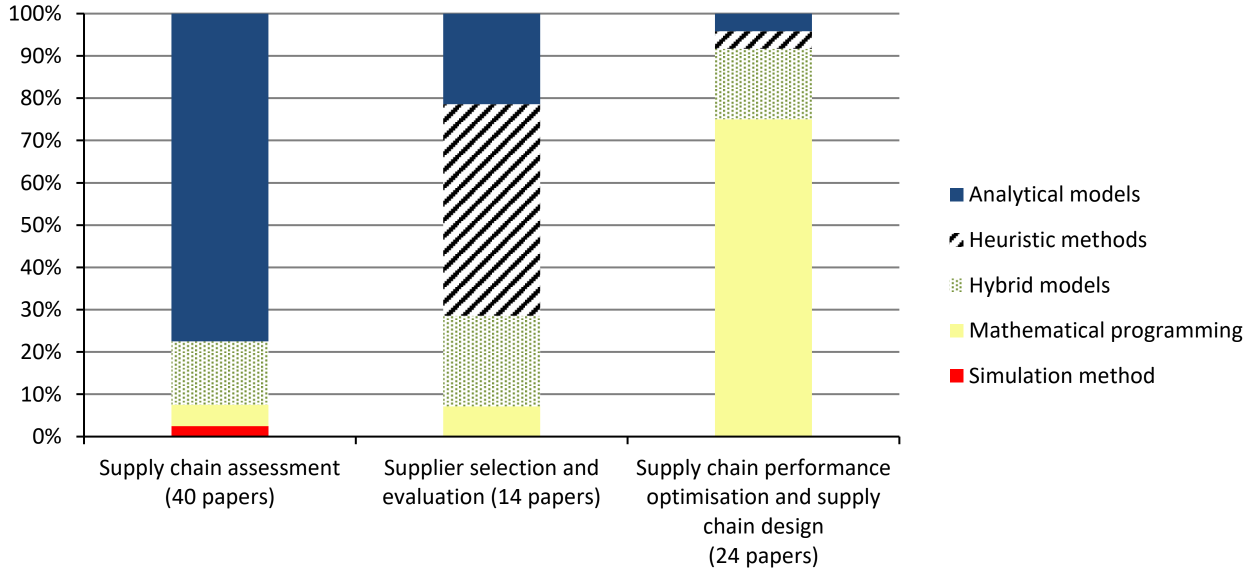

Fig. 9: Relationship between model type and scope of the methods

A contingency analysis was performed to support the findings from Figure 9 with a quantitative output by calculating Cramer's V. Once again, the single paper adopting simulation was excluded to ensure statistical validity of the test. Value of *t* is 2 in this case, therefore a large effect of association between variables is found for Cramer's V value above 0.35 (Cohen, 1988). Cramer's V was equal to 0.699 with a level of significance of 0.000, confirming a very strong association between model type and the scope of methods.

Synthesis of evaluated dimensions

Each scope includes unique characteristics beyond the method adopted, as depicted in Table VII, which are evaluated to synthesise all the dimensions considered in this review. Additionally, a discussion of the potential applicability by practitioners is made.

Supply chain design and performance optimisation papers provide a limited coverage of the supply chain: the majority is limited to short supply chains and typically adopts bi-objective optimisation including economic and environmental dimensions of sustainability. A key characteristic of these papers is the detailed supply chain modelling, which limits the extent of supply chain coverable by the performance measurement. Additionally, the environmental dimension shows a high prevalence in this type of papers: all papers, apart from Krikke (2010) and Manzardo et al., (2014) consider emissions to air, while other environmental inputs and outputs receive very limited attention. Natural resources and energy categories follow being represented in only 26% of the sample. As a result, supply chain design and performance optimisation papers generally tend to underestimate the overall supply chain environmental impact due to excessively narrow focus both in supply chain extent and environmental impacts. However, the dominant mathematical programming methods adopted support practitioners in operational improvement and decision making, by identifying optimal or near-optimal supply chain configurations in relation to the objective functions. These methods entail the determination of the physical structure of the supply chain as well as the flow of materials between tiers, enhancing the general visibility and traceability of the supply chain analysed.

Supplier selection and evaluation papers are by definition limited to only a dyad being involved in the measurement process. On the other hand, these methods balance the limited extent coverage with a wide range of environmental aspects, often including emissions to water, which are widely neglected in the literature, as pointed out in the "Environmental aspects" sub-section. Heuristic methods are the most prevalent model type for this scope, followed by analytical and hybrid models. Generally, these methods provide effective support for decision-making, either by ranking the suppliers, like in Sarkis and Dhavale (2015), or scoring the suppliers as well, like in Kannan et al. (2015).

Finally, papers focusing on assessment of the supply chain usually adopt analytical models, which are more frequently applied to extended supply chains as discussed earlier. Papers on "assessment of the supply chain" focus specifically on the measurement of the supply chain environmental performance, whereas the other two scope categories include this step as functional to other managerial decisions. Because of the combination between the methodological choices and the more specific focus, the supply chain extent covered by these papers is typically expanding beyond the 1st tier and showing applicability to extended supply chains in 40% of the instances. However, the extensive inclusion of supply chain tiers covered is accompanied by a narrower focus in terms of environmental inputs and outputs considered, limited to environmental inputs and emissions to air.

Table VII
Summary of the features of papers based on their primary scope

Paper scope	Environmental aspects	Extent of the supply chain	Dominant methodology
Supply chain design and performance optimisation	Limited scope Focus on air emissions	Not suitable for extended supply chains Various other supply chain configurations addressed	Mathematical programming
Supplier selection and evaluation	Complete evaluation of environmental inputs and outputs	Dyad supplier-focal firm	Heuristic methods dominant Hybrid models and systemic models also adopted
Assessment of the supply chain	Focus on resource consumption, energy and emissions to air	Multiple configurations of the supply chain measured: good applicability to extended supply chains	Analytical models

A trade-off can thus be identified between the extent of the supply chain and the range of the environmental aspects considered, indicating that a compromise is made. "Supplier evaluation and selection" methods perform best in environmental aspects range but are very limited in terms of supply chain extent, whereas "assessment of the supply chain" methods offer the best applicability to extended supply chains but consider limited environmental aspects.

6. Discussion

Implications for researchers and future research directions

Each research question led to a number of key findings and implications for researchers, while additional implications arose from the combined evaluation of research questions. These implications along with research directions are explicated in this section.

RQ1: What environmental performance metrics are adopted at the supply chain level?

A large variety of quantitative environmental measurements with very limited consistency was identified in the literature. Even though limiting the scope to environmental and quantitative measurements only, this finding confirms the analysis of Ahi and Searcy (2015) in the broader SSCM field. The growing body of literature on this topic is still at a divergent stage and a progressive standardisation in the future will be required to adopt similar units of reference. The extreme variety in the metrics adopted limits the applicability of developed methods for benchmarking applications. Environmental metrics are applied "to compare trends over time, to compare results with targets and to benchmark a company against others" (Gerbens-Leenes et al., 2003). While the first two objectives are achieved by the existing literature as consistency is achieved within the boundaries of each work, the last is currently missing due to the lack of standardisation in the metrics adopted and the lack of external reference values to compare results, thus making environmental measurements self-referential to specific studies and supply chains. Scholars often addressed the same environmental categories but adopted heterogeneous metrics, with very limited evidence of consideration of the metrics from techniques adopted by practitioners such as Global Reporting Initiative, SCOR model, Environmental European Agency or ISO 14000 series. Few exceptions include Mintcheva (2005), Nikolaou et al. (2013), Salvado et al. (2015) and Varsei et al., (2014). Therefore, this study calls for a standardisation of metrics in the GSCM future research to build a more homogenous body of research allowing the comparability of different studies and results. It also calls for a progressive merging of the perspectives from academia and industry in the future to further enhance the environmental metrics standardisation and studies comparability. While scholars can foster the development of standardised environmental metrics in the future, their application in operating contexts is largely dependent on the pressure companies are facing to adopt them. Regulatory bodies and third party organisations can effectively contribute towards the standardisation, whereas it is unlikely that this contribution will come from single supply chains as each is driven by different objectives.

The review also identified that a holistic evaluation of the environmental performance is still rare, with scholars focusing on limited sets of indicators that address specific environmental categories. Two patterns of environmental categories were identified thanks to contingency analysis. The efficiency oriented measurements tackling environmental aspects that generate monetary expenditure and the regulatory oriented measurements, which are largely based on the emissions to air. In the first case, interest for sustainable performance of supply chains is still led by the economic performance looking for win-win situations with the environmental performance, while in the latter case the regulatory schemes introduced in certain sectors and geographical areas triggered the interest of academics. Researchers will need to merge in future models these perspectives in order not only to obtain a holistic evaluation of the environmental performance but also to avoid a narrow approach to optimisation of the performance. Only the simultaneous consideration of all categories can lead to the identification of trade-offs between different environmental aspects and to a holistic improvement of the system examined.

RQ2: What extent of the supply chain, both upstream and downstream from the focal firm, are environmental performance measurement methods and related metrics addressing?

The findings show that attention is still limited to the 1st tier beyond the focal firm in the majority of cases, whereas the evaluation of extended supply chains is still at a developing stage. This finding highlights the need for improved supply chain traceability and visibility by the main players in the chain or the development of appropriate indirect mechanisms to reach sub-suppliers in multi-tier supply chains, to achieve a holistic supply chain-wide evaluation of the environmental performance.

The drawback of focusing on a limited supply chain extent appears particularly severe in the current competitive environment where global supply chains with multiple tiers are the norm (Kovács, 2008), as poor environmental performance of a single tier may cause an overall environmentally unsustainable behaviour of the entire supply chain (Miemczyk et al., 2012). The current dominant approach is paying attention only to direct business partners, demonstrating that GSCM is still far from being accomplished. The shift from green supplier selection to GSCM is still to be completed, at least for quantitative performance measurement of green supply chains; environmentally sustainable supply chains cannot be achieved by working only with first-tier partners (Genovese et al., 2013).

Research on quantitative models to measure supply chain environmental performance is still lagging behind in successfully reaching multi-tier and extended supply chain contexts. Therefore, this work calls for an expansion of the supply chain extent covered by GSCM performance measurement methods to achieve an effective supply chain-wide assessment and to avoid a potential underestimation of the true supply chain environmental impact. Identifying mechanisms to overcome the existing limited supply chain visibility and reach sub-suppliers located further upstream is a key challenge for researchers. Focal companies could either access directly the sub-suppliers augmenting their influence over the supply chain, work indirectly through their 1st tier suppliers and customers to access the extended supply chain or work with third parties (Tachizawa and Wong, 2014). While the direct and indirect approaches are driven by the organisations only, the last requires some supportive infrastructure. This could be provided either by NGOs, industry association or governmental bodies, which can pressure companies to address the environmental performance of their supply chain. Upcoming regulations such as the EU Environmental Footprint are an example. Working with external third parties would also contribute towards the standardisation of environmental metrics. Researchers will have to cope with challenges specific to the adoption of quantitative data across different organisations, including confidentiality and availability of data while taking into account the multiple organisation nature of supply chains. An interesting expansion to this work would be to look at the mechanisms adopted in multi-tier and extended supply chains to reach and collect data from sub-suppliers.

Finally, future research needs to pay particular attention to the downstream network, which is currently overlooked compared to the upstream network due to the limited liability of companies for the behaviour of their customers (Kovács, 2008). Measuring the environmental performance of usage and end-of-life management lifecycle stages looks critical, especially due to the complexity of accessing data (Michelsen et al., 2006). A key challenge for future research will be to develop methods to collect and share environmental data about product lifecycle stages that are beyond the control of any organisation in order to move from the dominant cradle-to-gate to the cradle-to-grave approach.

RQ3: What are the quantitative methods adopted to measure the environmental performance of supply chains? Is there a relationship between the type of method and the extent of supply chain covered or the scope of the work?

The analysis shows the dominance of two model types: mathematical programming and analytical models. Regarding the relationship between the type of method and the extent of supply chain covered, analytical models were identified by contingency analysis as the most frequently used to address extended supply chains, looking as the most promising method for future researchers to expand the supply chain extent coverage.

The paper scope and model types relationship exploration identified several strong associations: mathematical programming is primarily adopted for the design and optimisation of green supply chains, heuristic methods for green supplier selection and evaluation, and analytical models for the assessment of the supply chain performance. Considering the novelty of the research field, it is likely that the body of research will develop in three major streams in the future, based on a different purpose of the research and on consistent differences in the definition and boundaries of the supply chain. The analysis also revealed that papers focusing on the assessment of the supply chain show an excellent applicability to extended supply chains. On the other hand, supplier selection and evaluation papers proved to provide the most extensive coverage in terms of environmental aspects considered, for both environmental inputs and outputs.

RQ 1 & RQ 2: Environmental performance evaluation coverage and supply chain extent coverage

An identified key future challenge for researchers identified will be to overcome the observed trade-off between the scope of environmental performance and the extent of supply chain effectively measured. No paper analysed considers the extended supply chain while addressing and measuring all environmental aspects. The closest papers to this criterion are Koh et al. (2012), Michelsen et al. (2006) and Varsei et al. (2014), including four environmental categories while addressing extended supply chains and Adhitya et al., (2011), considering one upstream and two downstream tiers while still addressing all environmental categories.

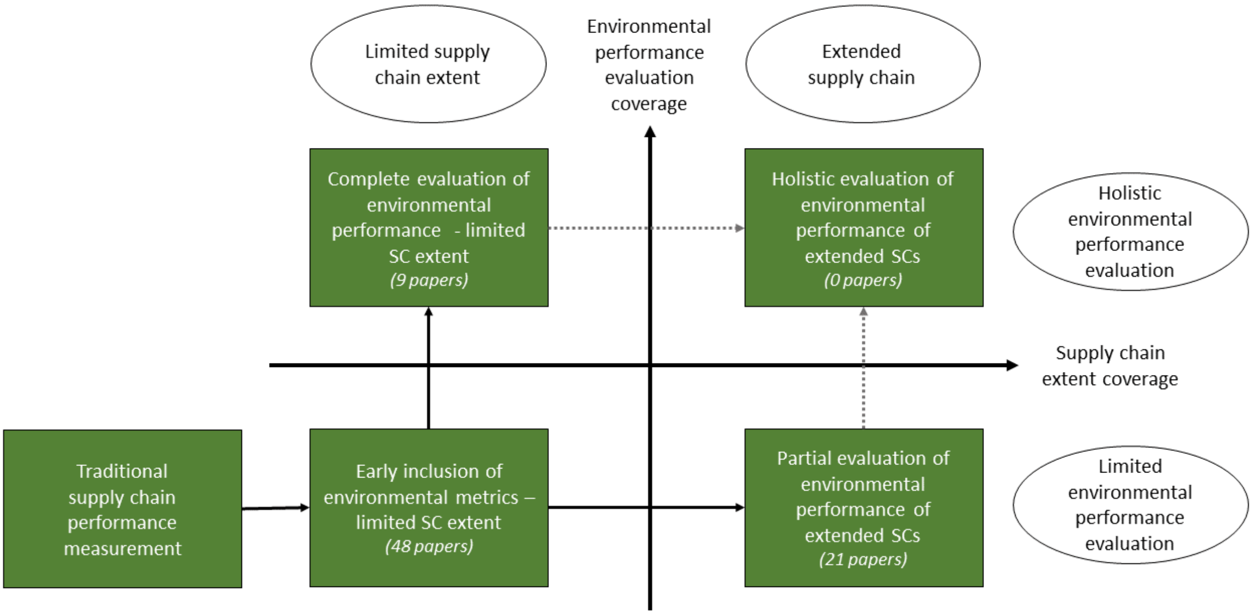

Fig. 10: Framework on the development of the GSCM performance measurement research field

Figure 10 presents a framework of the evolution of the literature up to now and the potential future directions. Supply chain performance measurement traditionally incorporated economic metrics along with other well-established key performance indicators such as time and quality (Beske-Janssen et al., 2015). The inclusion of environmental metrics followed as organisations recognised the importance of sustainability and of measuring non-financial aspects (Shen et al., 2013). However, environmental measurements were initially narrow in terms of scope, focusing on specific environmental categories while addressing a limited extent of the supply chain. Research further developed in two directions (solid arrows in Figure 10): either broadening the scope of environmental performance evaluation coverage or extending the supply chain extent coverage. However, no work was identified to progress sufficiently along both dimensions. Expanding simultaneously the comprehensiveness in terms of both environmental aspects considered and supply chain extent (dotted arrows in Figure 10) is identified as a key direction for future research. This would ultimately lead to a more comprehensive evaluation of the environmental performance of a supply chain, avoiding an underestimation of the true environmental impact due to too narrow approach in terms of environmental performance or extent of supply chain.

Limitations of the study

As every piece of research, this study is not immune from limitations, despite the emphasis in accuracy and rigour in the methodological choices made. First, the size and content of the sample is affected by the database selection: other databases apart from Scopus and Web of Science may have offered the opportunity to consider additional publications as well as the consideration of documents from the practitioner community. Moreover, as in every literature review process, a number of decisions still required a degree of subjectivity by reviewers, potentially affecting the final results. Finally, every type of classification, despite providing a structured and summarised understanding of the body of research, suffers from constraints and thus may not adequately convey the complexity and the specific in-depth features of every paper. This was particularly observed in the environmental measurements evaluation, where some measures fell among multiple categories and required an authors' decision about their classification, as well as in the supply chain extent evaluation, when the chain was described by the activities rather than the organisational entities involved. A careful analysis was required in these cases to assess the papers according to the categories adopted in the review.

Conclusions

This work aimed to identify quantitative methods developed to measure the environmental performance of supply chains, classify and evaluate their key features by systematically reviewing the literature at the intersection of the performance measurement and GSCM fields. 78 papers were evaluated according to the environmental aspects measured, the extent and type of supply chain addressed, the scope and the methods adopted. This work is the first to identify which supply chain tiers are effectively considered in environmental performance measurement along with the features of the supply chain examined. It also discussed the relationships between the supply chain characteristics and the other dimensions evaluated in the review. The analysis led to a number of key findings, leading to several implications for researchers.

A large variety of environmental measurements with very limited consistency was found in the literature. A progressive standardisation of environmental measurements to enhance benchmarking and the comparability of different studies would be beneficial in the future research. Moreover, the integration of the efficiency-oriented and the regulatory-oriented measurements would lead to a holistic evaluation of the environmental performance and avoid a narrow approach to the performance optimisation.

Regarding the supply chain extent coverage, the majority of methods do not measure the environmental performance beyond the 1st supply chain tier, while the evaluation of extended supply chains is still limited. Current approaches could strongly underestimate the true environmental impact of the supply chain by addressing the focal company and 1st tier business partners only, especially considering the global multi-tier nature of contemporary supply chains. Green supplier selection is a necessary step to improve the environmental performance, but it is only an intermediate step towards GSCM. The identification of mechanisms to reach sub-suppliers and to improve the visibility and traceability in multi-tier supply chains stand as key challenges for researchers. Finally, a gap in the measurement of downstream supply chains was identified, due to the challenges in accessing data from usage and end-of-life management stages.

Moreover, strong associations were identified both between the model types and the supply chain extent coverage and between the model types and the scope of the paper. Analytical models were identified as the most appropriate to measure the environmental performance of extended supply chains. The findings also suggest that the body of research of GSCM performance measurement is likely to develop along three main directions depending on the scope of the work, which will be complemented by consistent methodological choices and different definitions of the supply chain boundaries.

Finally, a trade-off between the environmental aspects and the supply chain extent coverage was highlighted, demonstrating that currently no method achieves a holistic evaluation of the supply chain environmental performance. Future research needs to address this gap in the literature and develop such methods and models, as highlighted in the framework of Figure 10.

The identification of several research gaps both through the three research questions and the combined evaluation of different dimensions of analysis represents the core contribution of this work, helping to shape future research and direct model development in the area of quantitative modelling for GSCM performance measurement.

This research is the first to analyse simultaneously the environmental metrics, the supply chain extent coverage and model types adopted for quantitative performance measurement of GSCM, exploring the relations between these three dimensions. The supply chain extent coverage is the most innovative aspect, providing a detailed mapping of what tiers of the supply chain are actually involved in environmental performance assessment. This helps to clarify the scope of the supply chain dimension in GSCM performance measurement research and to identify model types and other relevant features to expand the currently limited supply chain extent coverage.

Acknowledgments

This research was supported by a University of Strathclyde Research Studentship.

References

Adhitya, A., Halim, I. and Srinivasan, R. (2011), “Decision support for green supply chain operations by integrating dynamic simulation and LCA indicators: Diaper case study”, *Environmental Science and Technology*, Vol. 45 No. 23, pp. 10178–10185.

Ahi, P. and Searcy, C. (2013), “A comparative literature analysis of definitions for green and sustainable supply chain management”, *Journal of Cleaner Production*, Elsevier Ltd, Vol. 52, pp. 329–341.

Ahi, P. and Searcy, C. (2015), “An Analysis of Metrics Used to Measure Performance in Green and Sustainable Supply Chains”, *Journal of Cleaner Production*, Elsevier Ltd, Vol. 86, pp. 360–377.

APICS. (2012), *APICS 2012 Sustainability Challenges and Practices*.

APICS. (2014), *SCOR*, available at:https://doi.org/10.1007/s13398-014-0173-7.2.

Beavis, L. (2015), “M&S takes an interest in its suppliers’ green credentials”, *The Guardian*, available at: https://www.theguardian.com/sustainable-business/2015/apr/30/ms-takes-an-interest-in-its-suppliers-green-credentials (accessed 22 December 2016).

Bell, J.E., Mollenkopf, D.A. and Stolze, H.J. (2013), "Natural resource scarcity and the closed- loop supply chain: a resource-advantage view", *International Journal of Physical Distribution & Logistics Management*, Vol. 43 No. 5/6, pp. 351–379.

Beske-Janssen, P., Johnson, M.P. and Schaltegger, S. (2015), "20 Years of Performance Measurement in Sustainable Supply Chain Management – What Has Been Achieved?", *Supply Chain Management: An International Journal*, Vol. 20 No. 6, pp. 664–680.

Björklund, M., Martinsen, U. and Abrahamsson, M. (2012), "Performance measurements in the greening of supply chains", *Supply Chain Management: An International Journal*, Vol. 17 No. 1, pp. 29–39.

Bloemhof, J.M., van der Vorst, J.G. a. J., Bastl, M. and Allaoui, H. (2015), "Sustainability assessment of food chain logistics", *International Journal of Logistics Research and Applications*, Vol. 18 No. 2, pp. 101–117.

Bojarski, A.D., Laínez, J.M., Espuña, A. and Puigjaner, L. (2009), "Incorporating environmental impacts and regulations in a holistic supply chains modeling: An LCA approach", *Computers and Chemical Engineering*, Vol. 33 No. 10, pp. 1747–1759.

Bouchery, Y., Ghaffari, A., Jemai, Z. and Dallery, Y. (2012), "Including sustainability criteria into inventory models", *European Journal of Operational Research*, Elsevier B.V., Vol. 222 No. 2, pp. 229–240.

Boukherroub, T., Ruiz, A., Guinet, A. and Fondrevelle, J. (2014), "An integrated approach for sustainable supply chain planning", *Computers & Operations Research*, Elsevier, Vol. 54, pp. 180–194.

Brandenburg, M., Govindan, K., Sarkis, J. and Seuring, S. (2014), "Quantitative models for sustainable supply chain management: Developments and directions", *European Journal of Operational Research*, Elsevier B.V., Vol. 233 No. 2, pp. 299–312.

Brent, a. C. and Visser, J.K. (2005), "An environmental performance resource impact indicator for life cycle management in the manufacturing industry", *Journal of Cleaner Production*, Vol. 13 No. 6, pp. 557–565.

Cabral, I., Grilo, A. and Cruz-Machado, V. (2012), "A decision-making model for Lean, Agile, Resilient and Green supply chain management", *International Journal of Production Research*, Vol. 50 No. 17, pp. 4830–4845.

Caro, F., Corbett, C.J., Tan, T. and Zuidwijk, R. (2013), "Double Counting in Supply Chain Carbon Footprinting", *Manufacturing & Service Operations Management*, Vol. 15 No. 4, pp. 545–558.

Chatzinikolaou, S.D. and Ventikos, N.P. (2015), "Holistic framework for studying ship air emissions in a life cycle perspective", *Ocean Engineering*, Elsevier, Vol. 110, pp. 113–122.

Cichorowski, G., Joa, B., Hottenroth, H. and Schmidt, M. (2015), "Scenario analysis of life cycle greenhouse gas emissions of Darjeeling tea", *The International Journal of Life Cycle Assessment*, Vol. 20 No. 4, pp. 426–439.

Cohen, J. (1969), *Statistical Power Analysis for the Behavioural Sciences*, Academic, London.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum, Hillsdale, New Jersey.

De Soete, W., Dewulf, J., Cappuyns, P., Van der Vorst, G., Heirman, B., Aelterman, W., Schoeters, K., et al. (2013), "Exergetic sustainability assessment of batch versus continuous wet granulation based pharmaceutical tablet manufacturing: a cohesive analysis at three different levels", *Green Chemistry*, Vol. 15 No. 11, p. 3039.

Dotoli, M., Fanti, M.P., Meloni, C. and Zhou, M.C. (2006), “Design and optimization of integrated e-supply chain for agile and environmentally conscious manufacturing”, *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, Vol. 36 No. 1, pp. 62–75.

Elkington, J. (2004), “Enter the Triple Bottom Line”, *The Triple Bottom Line: Does It All Add Up?*, Vol. 1, Earthscan, New York, pp. 1–16.

Erol, I., Sencer, S. and Sari, R. (2011), “A new fuzzy multi-criteria framework for measuring sustainability performance of a supply chain”, *Ecological Economics*, Vol. 70 No. 6, pp. 1088–1100.

Fabbe-Costes, N., Roussat, C. and Colin, J. (2011), “Future sustainable supply chains: what should companies scan?”, *International Journal of Physical Distribution & Logistics Management*, Vol. 41 No. 3, pp. 228–252.

Fahimnia, B., Sarkis, J. and Eshragh, A. (2015), “A tradeoff model for green supply chain planning: A leanness-versus-greenness analysis”, *Omega*, Elsevier, Vol. 54, pp. 173–190.

Fink, A. (1998), *Conducting Research Literature Reviews: From Paper to the Internet*, Sage, Thousand Oaks.

Fleiss, J.L. (1981), *Statistical Methods for Rates and Proportions*, Wiley, New York.

Genovese, A., Lenny Koh, S.C., Kumar, N. and Tripathi, P.K. (2013), “Exploring the challenges in implementing supplier environmental performance measurement models: a case study”, *Production Planning & Control*, Vol. 7287 No. July 2014, pp. 1–14.

Gerbens-Leenes, P.W., Moll, H.C. and Schoot Uiterkamp, a. J.M. (2003), “Design and development of a measuring method for environmental sustainability in food production systems”, *Ecological Economics*, Vol. 46 No. 2, pp. 231–248.

Gimenez, C. and Tachizawa, E.M. (2012), “Extending sustainability to suppliers: a systematic literature review”, *Supply Chain Management: An International Journal*, Vol. 17 No. 5, pp. 531–543.

Gold, S., Seuring, S. and Beske, P. (2010), “Sustainable supply chain management and inter-organizational resources: A literature review”, *Corporate Social Responsibility and Environmental Management*, Vol. 17 No. 4, pp. 230–245.

Hervani, A.A., Helms, M.M. and Sarkis, J. (2005), “Performance measurement for green supply chain management”, *Benchmarking: An International Journal*, Vol. 12 No. 4, pp. 330–353.

Hutchins, M.J. and Sutherland, J.W. (2008), “An exploration of measures of social sustainability and their application to supply chain decisions”, *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1688–1698.

Jamshidi, R., Fatemi Ghomi, S.M.T. and Karimi, B. (2012), “Multi-objective green supply chain optimization with a new hybrid memetic algorithm using the Taguchi method”, *Scientia Iranica*, Elsevier B.V., Vol. 19 No. 6, pp. 1876–1886.

Jesson, J.K., Matheson, L. and Lacey, F.M. (2011), *Doing Your Literature Review*, Sage, London.

Kannan, D., Govindan, K. and Rajendran, S. (2015), “Fuzzy axiomatic design approach based green supplier selection: A case study from Singapore”, *Journal of Cleaner Production*, Elsevier Ltd, Vol. 96, pp. 194–208.

Kateri, M. (2014), *Contingency Table Analysis: Methods and Implementation Using R*, Springer, New York, available at:https://doi.org/10.1002/0470011815.b2a10016.

Koh, S.C.L., Genovese, A., Acquaye, A. a., Barratt, P., Rana, N., Kuylenstierna, J. and Gibbs, D. (2012), “Decarbonising product supply chains: design and development of an integrated evidence-based decision support system – the supply chain environmental analysis tool (SCEnAT)", *International Journal of Production Research*, Vol. 51 No. 7, pp. 1–18.

Kovács, G. (2008), "Corporate environmental responsibility in the supply chain", *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1571–1578.

Krikke, H. (2010), "Opportunistic versus life-cycle-oriented decision making in multi-loop recovery: An eco-eco study on disposed vehicles", *International Journal of Life Cycle Assessment*, Vol. 15, pp. 757–768.

Krikke, H. (2011), "Impact of closed-loop network configurations on carbon footprints: A case study in copiers", *Resources, Conservation and Recycling*, Elsevier B.V., Vol. 55 No. 12, pp. 1196–1205.

Liebtrau, A.M. (1983), *Measures of Association*, Sage, Newbury Park, CA, USA.

Liu, S., Leat, M. and Smith, M.H. (2011), "State-of-the-art sustainability analysis methodologies for efficient decision support in green production operations", *International Journal of Sustainable Engineering*, Vol. 4 No. August 2015, pp. 236–250.

Manzardo, A., Ren, J., Piantella, A., Mazzi, A., Fedele, A. and Scipioni, A. (2014), "Integration of water footprint accounting and costs for optimal chemical pulp supply mix in paper industry", *Journal of Cleaner Production*, Elsevier Ltd, Vol. 72, pp. 167–173.

McIntyre, K., Smith, H., Henham, a and Pretlove, J. (1998), "Environmental performance indicators for integrated supply chains: The case of Xerox Ltd", *Supply Chain Management*, Vol. 3 No. 3, pp. 149–156.

Meixell, M.J. and Luoma, P. (2015), "Stakeholder pressure in sustainable supply chain management", *International Journal of Physical Distribution & Logistics Management*, Vol. 45 No. 1/2, pp. 69–89.

Mellor, W., Wright, E., Clift, R., Azapagic, A. and Stevens, G. (2002), "A mathematical model and decision-support framework for material recovery, recycling and cascaded use", *Chemical Engineering Science*, Vol. 57 No. 22–23, pp. 4697–4713.

Mentzer, J.T., Keebler, J.S., Nix, N.W., Smith, C.D. and Zacharia, Z.G. (2001), "Defining Supply Chain Management", *Journal of Business Logistics*, Vol. 22 No. 2, pp. 1–25.

Michelsen, O. and Fet, A.M. (2010), "Using eco-efficiency in sustainable supply chain management; A case study of furniture production", *Clean Technologies and Environmental Policy*, Vol. 12 No. 5, pp. 561–570.

Michelsen, O., Fet, A.M. and Dahlsrud, A. (2006), "Eco-efficiency in extended supply chains: A case study of furniture production", *Journal of Environmental Management*, Vol. 79 No. 3, pp. 290–297.

Miemczyk, J., Johnsen, T.E. and Macquet, M. (2012), "Sustainable purchasing and supply management: a structured literature review of definitions and measures at the dyad, chain and network levels", *Supply Chain Management: An International Journal*, Vol. 17 No. 5, pp. 478–496.

Mintcheva, V. (2005), "Indicators for environmental policy integration in the food supply chain (the case of the tomato ketchup supply chain and the integrated product policy)", *Journal of Cleaner Production*, Vol. 13 No. 7, pp. 717–731.

Nagel, M.H. (2003), "Managing the environmental performance of production facilities in the electronics industry: More than application of the concept of cleaner production", *Journal of Cleaner Production*, Vol. 11 No. 1, pp. 11–26.

Neely, A., Gregory, M. and Platts, K. (1995), "Performance measurement system design: A literature review and research agenda", *International Journal of Operations & Production Management*, Vol. 15 No. 4, pp. 80–116.

Nikolaou, I.E., Evangelinos, K.I. and Allan, S. (2013), "A reverse logistics social responsibility evaluation framework based on the triple bottom line approach", *Journal of Cleaner Production*, Elsevier Ltd, Vol. 56, pp. 173–184.

O'Rourke, D. (2014), "The science of sustainable supply chains", *Science*, Vol. 344 No. 6188, pp. 1124–1128.

Pålsson, H., Finnsgård, C. and Wänström, C. (2013), "Selection of Packaging Systems in Supply Chains from a Sustainability Perspective: The Case of Volvo", *Packaging and Technology and Science*, Vol. 26, pp. 289–310.

Pattara, C., Raggi, A. and Cichelli, A. (2012), "Life cycle assessment and carbon footprint in the wine supply-chain", *Environmental Management*, Vol. 49 No. 6, pp. 1247–1258.

Rao, P. and Holt, D. (2005), "Do green supply chains lead to competitiveness and economic performance?", *International Journal of Operations & Production Management*, Vol. 25 No. 9, pp. 898–916.

Ren, J., Tan, S., Yang, L., Goodsite, M.E., Pang, C. and Dong, L. (2015), "Optimization of emergy sustainability index for biodiesel supply network design", *Energy Conversion and Management*, Elsevier Ltd, Vol. 92, pp. 312–321.

Renewable Choice Energy. (2012), *The Growing Trend of Sustainability Scorecards*.

Ritthof, M., Rohn, H., Liedtke, C. and Merten, T. (2002), *Calculating MIPS Resource Productivity of Products and Services*, Wuppertal Institut for Climate, Environment and Energy at the Science Centre North Rhine-Westphalia.

Salvado, M., Azevedo, S., Matias, J. and Ferreira, L. (2015), "Proposal of a Sustainability Index for the Automotive Industry", *Sustainability*, Vol. 7 No. 2, pp. 2113–2144.

Sarkis, J. and Dhavale, D.G. (2015), "Supplier selection for sustainable operations: A triple-bottom-line approach using a Bayesian framework", *International Journal of Production Economics*, Elsevier, Vol. 166, pp. 177–191.

Saunders, M., Thornhill, A. and Lewis, P. (1997), *Research Methods for Business Students*, Paerson Education, Harlow.

Schmidt, M. and Schwegler, R. (2008), "A recursive ecological indicator system for the supply chain of a company", *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1658–1664.

Seuring, S. (2013), "A review of modeling approaches for sustainable supply chain management", *Decision Support Systems*, Elsevier B.V., Vol. 54 No. 4, pp. 1513–1520.

Seuring, S. and Müller, M. (2008), "From a literature review to a conceptual framework for sustainable supply chain management", *Journal of Cleaner Production*, Vol. 16 No. 15, pp. 1699–1710.

Shaw, S., Grant, D.B. and Mangan, J. (2010), "Developing environmental supply chain performance measures", *Benchmarking: An International Journal*, Vol. 17 No. 3, pp. 320–339.

Shen, L., Olfat, L., Govindan, K., Khodaverdi, R. and Diabat, A. (2013), "A fuzzy multi criteria approach for evaluating green supplier's performance in green supply chain with linguistic preferences", *Resources, Conservation and Recycling*, Elsevier B.V., Vol. 74, pp. 170–179.

Slack, N., Chambers, S., Johnston, R. and Betts, A. (2009), *Operations and Process Management*, Paerson Education, Harlow.

Srivastava, S.K. (2007), "Green supply-chain management: A state-of-the-art literature review", *International Journal of Management Reviews*, Vol. 9 No. 1, pp. 53–80.

Stevens, G.C. (1989), “Integrating the Supply Chain”, *International Journal of Physical Distribution & Materials Management*, Vol. 19 No. 8, pp. 3–8.

Tachizawa, E.M. and Wong, C.Y. (2014), “Towards a theory of multi-tier sustainable supply chains: a systematic literature review”, *Supply Chain Management: An International Journal*, Vol. 19 No. 5/6, pp. 643–663.

Taticchi, P., Garengo, P., Nudurupati, S.S., Tonelli, F. and Pasqualino, R. (2014), “A review of decision-support tools and performance measurement and sustainable supply chain management”, *International Journal of Production Research*, Vol. 53 No. 21, pp. 1–22.

Taticchi, P., Tonelli, F. and Pasqualino, R. (2013), “Performance measurement of sustainable supply chains: A literature review and a research agenda”, *International Journal of Productivity and Performance Management*, Vol. 62 No. 8, pp. 782–804.

Tranfield, D., Denyer, D. and Smart, P. (2003), “Towards a methodology for developing evidence-informed management knowledge by means of systematic review”, *British Journal of Management*, Vol. 14, pp. 207–222.

Trappey, A.J.C., Trappey, C. V., Hsiao, C.-T., Ou, J.J.R. and Chang, C.-T. (2012), “System dynamics modelling of product carbon footprint life cycles for collaborative green supply chains”, *International Journal of Computer Integrated Manufacturing*, Vol. 25 No. 10, pp. 934–945.

UNEP. (2010), *Assessing the Environmental Impacts of Consumption and Production. Priority Products and Materials*, *A Report of the Working Group on the Environmental Impacts of Products and Materials to the International Panel for Sustainable Resource Management*.

Varsei, M., Soosay, C., Fahimnia, B. and Sarkis, J. (2014), “Framing sustainability performance of supply chains with multidimensional indicators”, *Supply Chain Management: An International Journal*, Vol. 19 No. 3, pp. 242–257.

Veleva, V., Hart, M., Greiner, T. and Crumbley, C. (2003), “Indicators for measuring environmental sustainability: A case study of the pharmaceutical industry”, *Benchmarking: An International Journal*, Vol. 10 No. 2, pp. 107–119.

WBCSD and WRI. (2009), *The Greenhouse Gas Protocol Initiative: Scope 3 Accounting and Reporting Standard*, Geneva, Switzerland.

Zakeri, A., Dehghanian, F., Fahimnia, B. and Sarkis, J. (2015), “Carbon pricing versus emissions trading: A supply chain planning perspective”, *International Journal of Production Economics*, Elsevier, Vol. 164, pp. 197–205.

Zhang, Q., Shah, N., Wassick, J., Helling, R. and van Egerschot, P. (2014), “Sustainable supply chain optimisation: An industrial case study”, *Computers & Industrial Engineering*, Elsevier Ltd, Vol. 74, pp. 68–83.